Oak Ridge National Lab Update on Cray XT3

NATIONAL CENTER
FOR COMPUTATIONAL SCIENCES

presented by

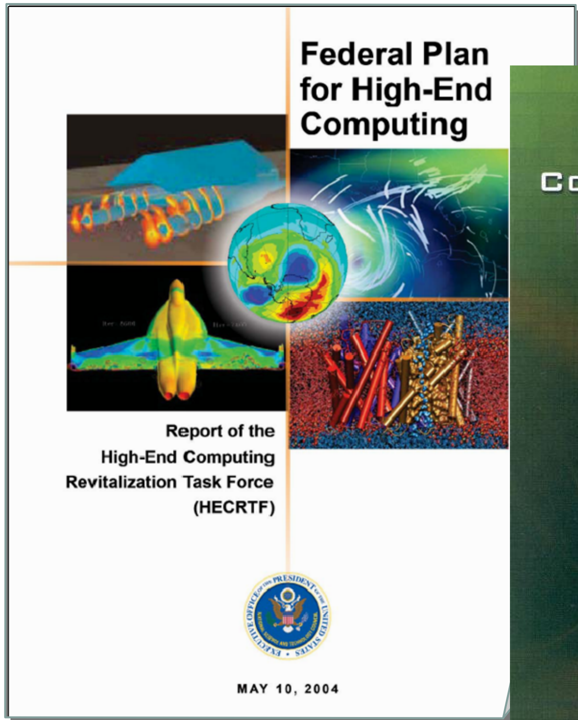
Sarp Oral, Ph.D.

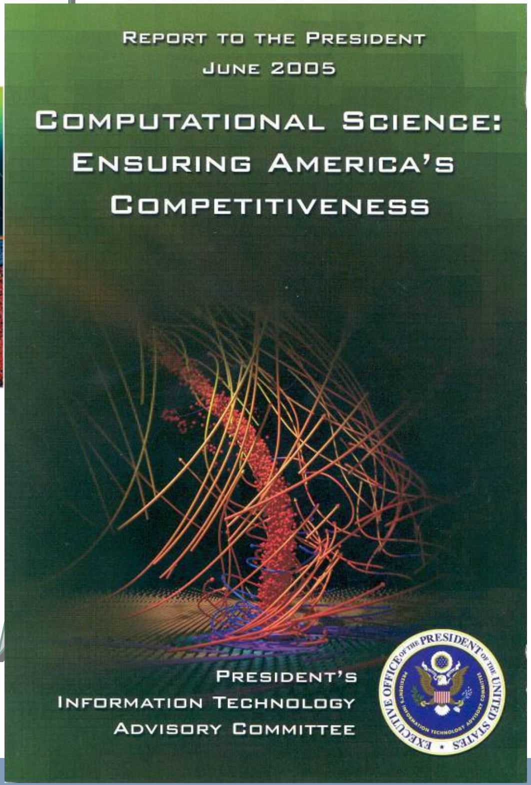
7th LCI International Conference on Clusters, May 3rd 2006

Oak Ridge National Laboratory
U.S. Department of Energy

Leadership Computing is a National Priority

"The goal of such systems [leadership systems] is to provide computational capability that is at least 100 times greater than what is currently available."





"High-end system deployments should be viewed not as an interagency competition but as a shared strategic need that requires coordinated agency responses."

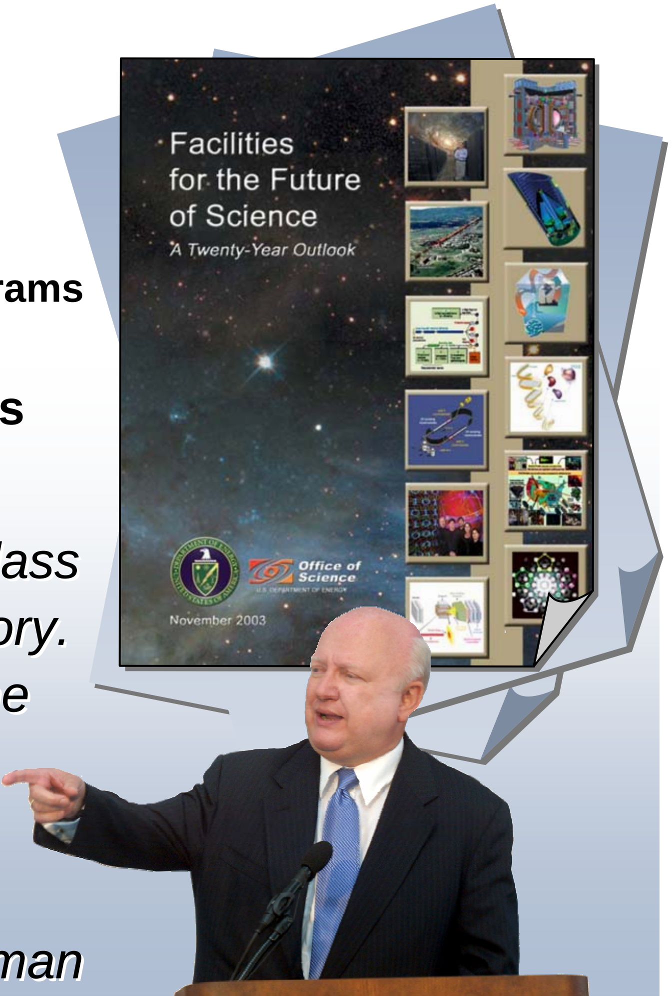
In 2004 ORNL's NCCS was selected as the
National Leadership Computing Facility

Leadership Computing is the Highest Domestic Priority of the Office of Science

- **Ray Orbach has articulated his philosophy for the SC laboratories**
  - **Each lab will have world-class capabilities in one or more areas of importance to Office of Science**
  - **ORNL: SNS and NCCS will underpin world-class programs in materials, energy, and life sciences**
- **20-year facilities plan being used to set priorities among projects**

"I am committed to the concept of a Leadership Class Computing facility at Oak Ridge National Laboratory. The facility will be used to meet the missions of the Department and those of other agencies. I can assure you that I understand the important role supercomputing plays in scientific discovery."

Secretary Bodman



NCCS Mission to Enable Science Success

World leader in scientific computing

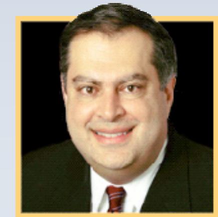
"User facility providing leadership-class computing capability to scientists and engineers nationwide independent of their institutional affiliation or source of funding"

Intellectual center in computational science

Create an interdisciplinary environment where science and technology leaders converge to offer solutions to tomorrow's challenges

Transform scientific discovery through advanced computing

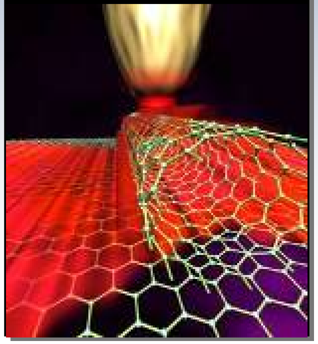
"Deliver major research breakthroughs, significant technological innovations, medical and health advances, enhanced economic competitiveness, and improved quality of life for the American people"



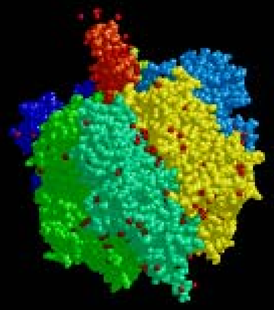
– Secretary Abraham

Key National Science Priorities

Manipulating the Nanoworld

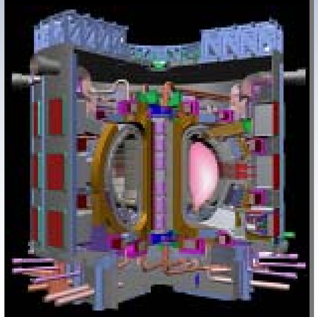


Taming the Microbial World

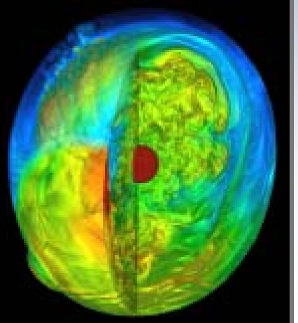


Environment and Health

ITER for Fusion Energy



Search for the Beginning

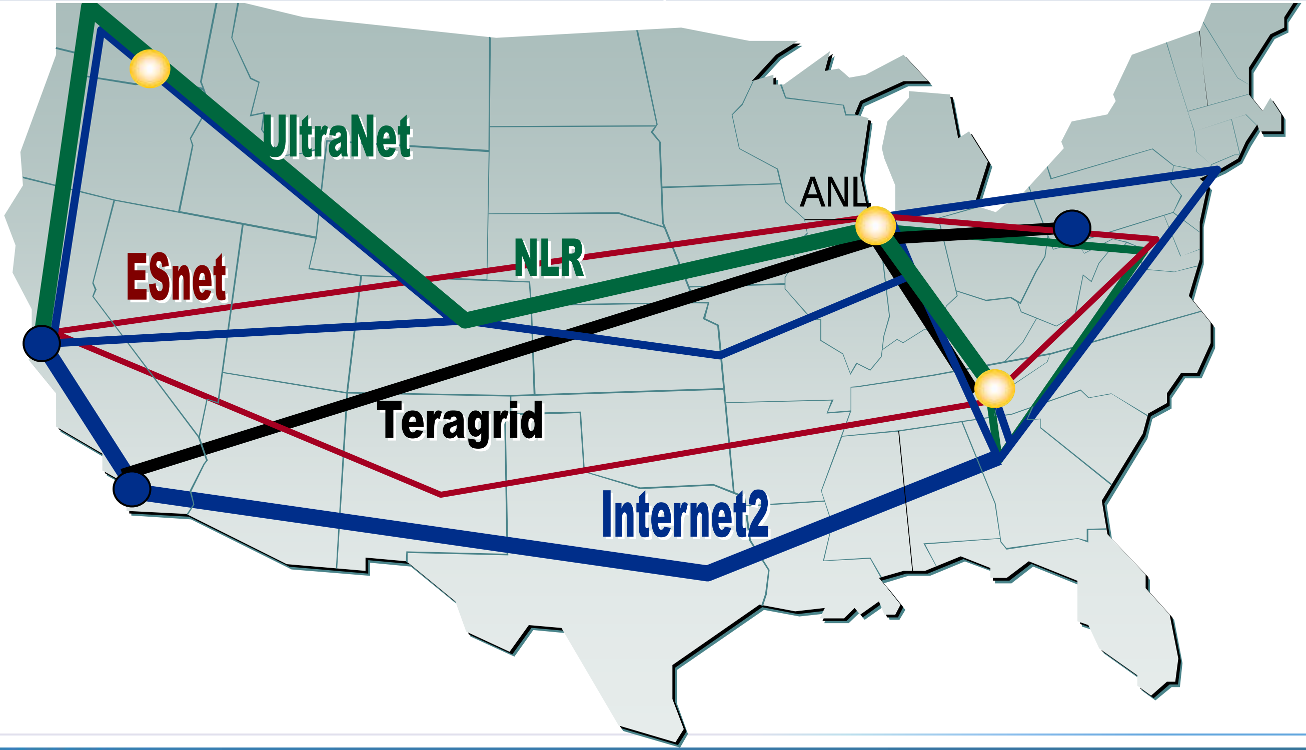


Recent NCCS research includes:

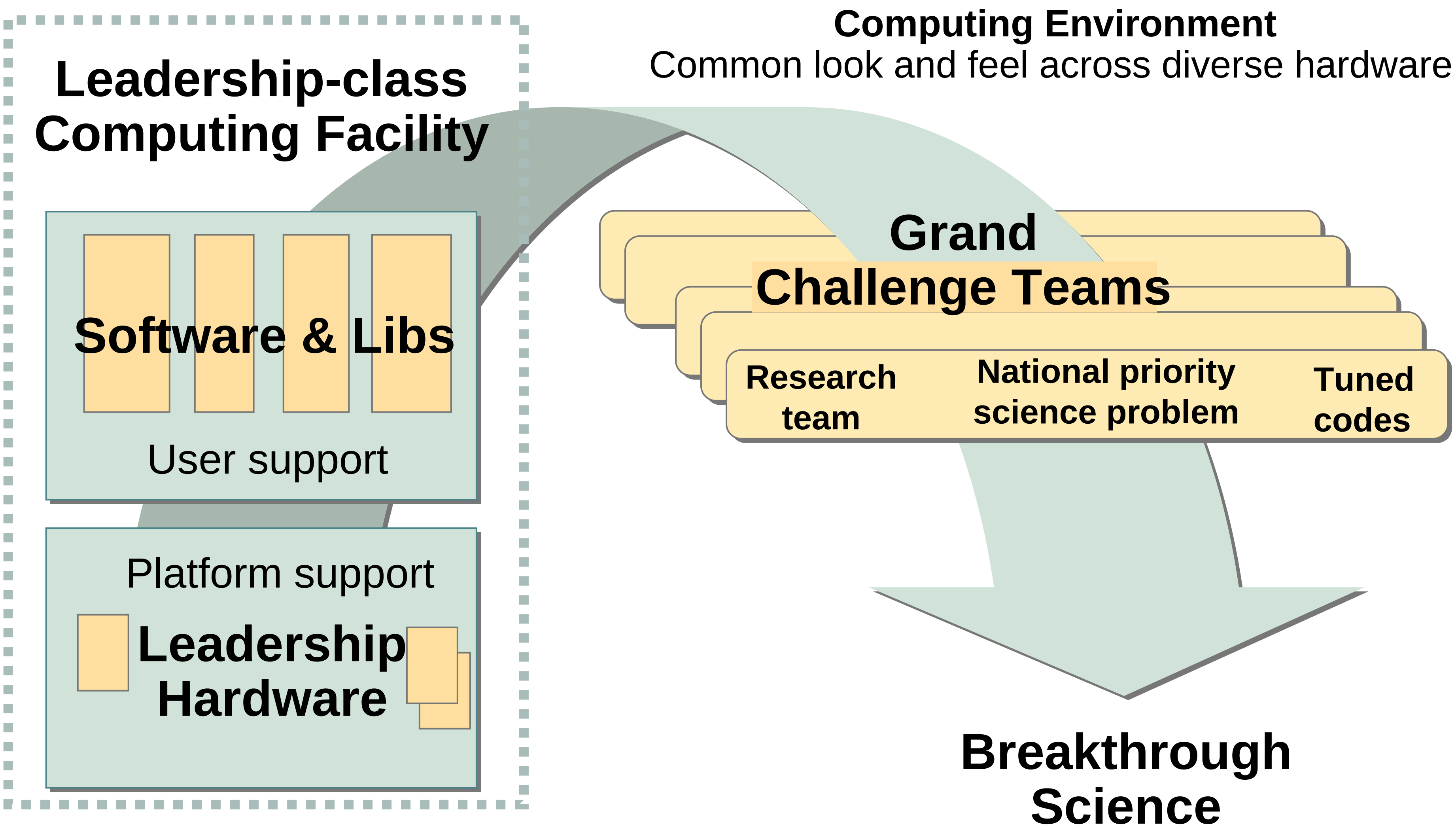
Largest simulation of plasma behavior in a tokamak
Resolution of theoretical disputes in materials research
Identification of shock wave instability in supernovae collapse
Seeing interplay of complex chemistry in combustion

High Bandwidth Connectivity to NLCF Enable Efficient Remote User Access

Connected to Major Science Networks	
OC48 to ESNET (provisioned by ESNET)	1 - 4 x 10 Gb to NSF Teragrid
10 Gb to Internet2	2 x 10 Gb Ultranet
2 x 10 Gb to National Lambda Rail	12 x 10 Gb Futurenet



NCCS Environment



Project Types

1. Grand Challenge

- Scientific problems that may only be addressed through access to NCCS hardware, software, and science expertise
- Multi-year, multi-million CPU hour allocation

2. Pilot Project

- Small allocations for projects, in preparation for future Grand Challenge or End Station submittals
- Limited in duration

3. End Station

Computationally intense research projects, also dedicated to development of community applications

Access to NLCF

Proposals

- Pilot Projec

NATIONAL CENTER FOR COMPUTATIONAL SCIENCES

Review

- Technical readin… s
- Scalability

Allocations

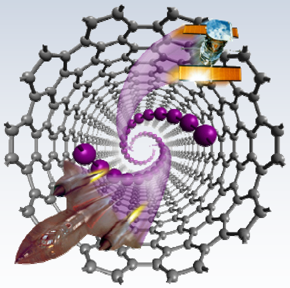
- Grand Challenge
- End Stations
- Pilot Projects

Computational End Station

NCCS deploys a fundamentally new approach for long-term engagement of research communities modeled on the "end station" concept through which major experimental facilities provide specialized instruments to specific user groups

End Station defined by three characteristics:

1 **Addresses problems that are of national importance (e.g., nanotech)**



National Problem

+

2 **Scientific team willing to create and maintain end station**

Scientific Team

+

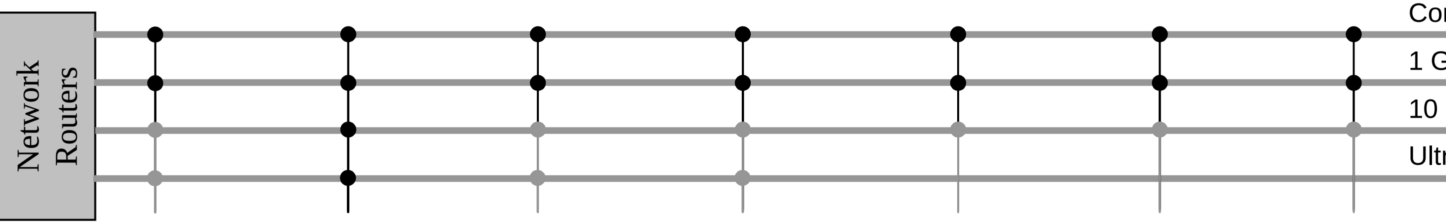
3 **Suite of scientific codes in area tuned to NCCS resources**

Application Suite

Scientific Needs Push Leadership-Class Computing to the Edge

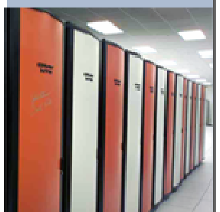
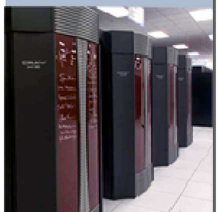
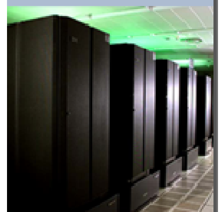
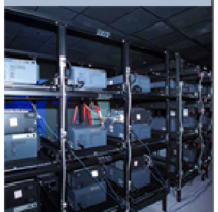
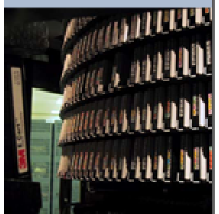
- NEW MATERIAL DESIGN
- CLIMATE CHANGE POLICY
- ALTERNATIVE ENERGY SOURCES
- MOLECULAR CONTROL OF CELLULAR PROCESSES

NCCS Resources



September 2005 Summary

7 Systems

Cray XT3 Jaguar	Cray X1E Phoenix	SGI Altix Ram	IBM SP4 Cheetah	IBM Linux NSTG	Visualization Cluster	IBM HPSS
						
(5,294) 2.4GHz 11TB Memory	**(1,024) 0.5GHz 2 TB Memory**	**(256) 1.5GHz 2TB Memory**	**(864) 1.3GHz 1.1TB Memory**	**(56) 3GHz 76GB Memory**	**(128) 2.2GHz 128GB Memory**	**Many Storage Devices Supported**
120TB	**32TB**	**36TB**	**32TB**	**4.5TB**	**9TB**	**5TB**

Supercomputers
7,622 CPUs
16TB Memory
45 TFlops

Total Shared Disk
238.5 TB

Scientific Visualization Lab

27 projector,
35 megapixel,
Power Wall

Test Systems

- 96 processor Cray XT3
- 32 processor Cray X1E*
- 16 Processor SGI Altix

Evaluation Platforms

- 144 processor Cray XD1 with FPGAs
- SRC Mapstation
- Clearspeed
- BlueGene (at ANL)

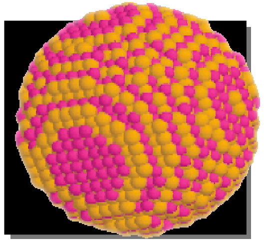
Backup Storage

5PB

5 PB

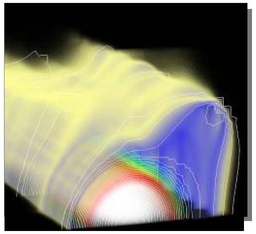
Jaguar

Accepted in 2005 and routinely running applications requiring 4,000 to 5,000 processors



Materials Science

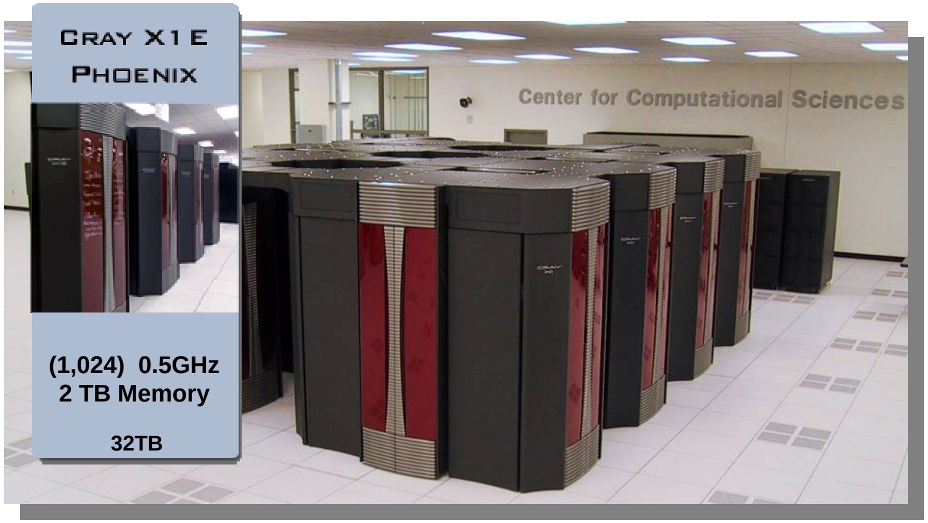
Nanoparticles present the capacity for information storage dramatically greater than bulk materials. Over 81% of theoretical peak performance was achieved for the non-collinear magnetic structure calculation of FePt particles.



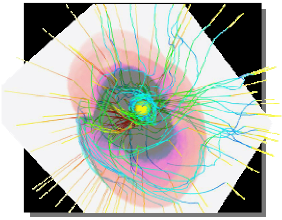
Plasma Turbulence

Largest-ever simulation of plasma behavior in a tokamak crucial to harness the power of fusion reactions; simulation used 60% of Jaguar resources.

Phoenix

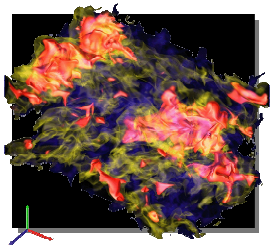


Highly scalable hardware and software
High sustained performance on real applications



Astrophysics

Simulations have uncovered a new instability of the shock wave and a resultant spin-up of the stellar core beneath it, which may explain key observables such as neutron star “kicks” and the spin of newly-born pulsars.



Combustion

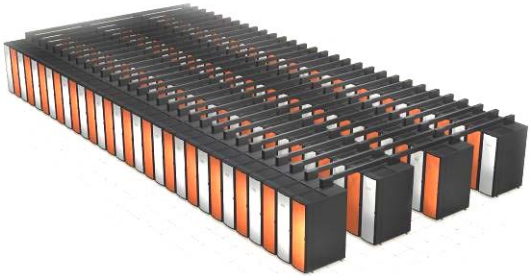
Calculations show the importance of the interplay of diffusion and reaction, particularly where strong finite-rate chemistry effects are involved.

Jaguar System Overview

- 10^{th} fastest supercomputer
  - **Top500 list**
  - **25 TFLOPS @ 5200 processors**
- Total 5294 nodes
  - **Single core AMD Opterons @ 2.4 GHz**
  - **5212 compute nodes**
  - **82 service nodes**
    - **Network, login, I/O etc.**
- 3-D torus topology
- UNICOS/ls OS
  - **Linux on service nodes**
  - **Catamount on compute nodes**
- Maximum up time
  - **Currently 7 days**
- Maximum utilization
  - **Currently ~ 85%**

Jaguar High-Performance File System

- Lustre
  - **High-performance scratch file system**
- Current configuration
  - **24 OSS and 48 OSTs on 6 DDN 8500 couplets**
  - **~ 38.5 TB disk space**
- Final configuration
  - **48 OSS and 64 OST and 14 DDN 8500 couplets**
  - **~ 96 TB disk space**
- Maximum I/O performance
  - **~ 7.5 GB/s R/W single shared file**
    - **64 OST on 32 OSS**
    - **128 clients**



NCCS Infrastructure Systems

High Performance Storage System

Multi-Petabyte data archive used by HPC centers around the world
Developed by ORNL, LLNL, LANL, SNL, LBNL, and IBM

Data Analysis and Visualization Cluster

128 AMD Opteron processors
Quadrics Interconnect

Visualization Facility

27’ x 8’ Display wall with 35 megapixels
IDesk, Cave, 9 megapixel display
Chromium, AVS, Remote visualization software

NCCS Software Infrastructure

Operating Systems

Linux on all new systems
Unix variants (AIX, Unicos/MP)
Batch systems: Loadleveler, PBSpro

File Systems Strategy

Unified home directories on NFS
Local scratch file systems
High speed parallel file systems (Lustre, GPFS)
HPSS Archival storage

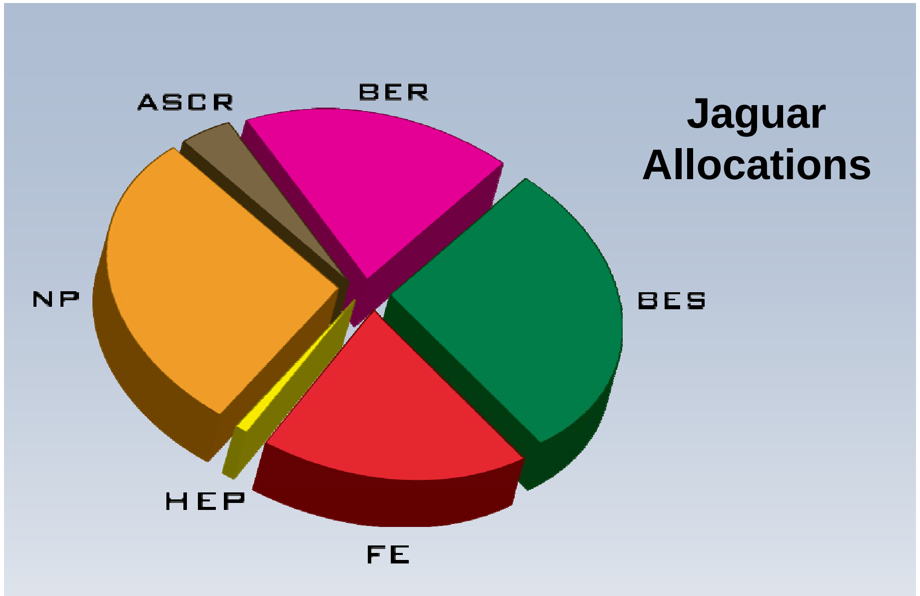
Programming Environment

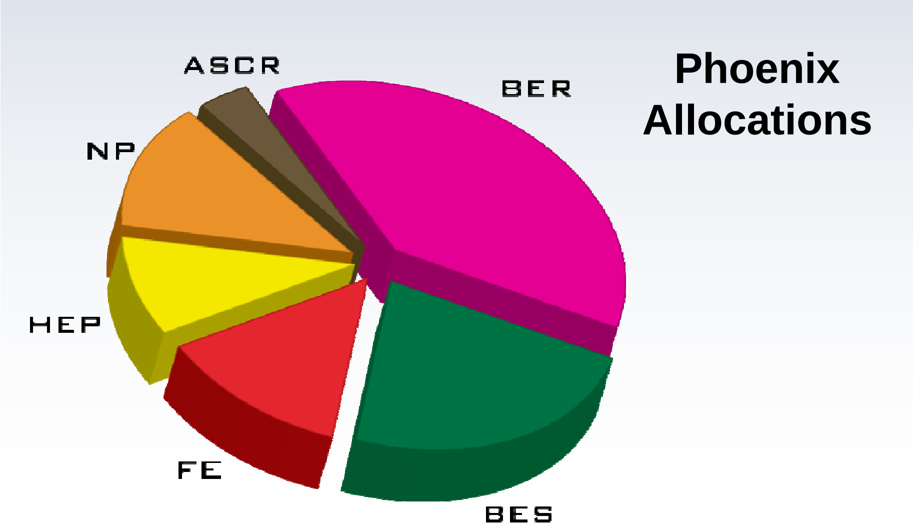
Fortran, C, C++, Coarray Fortran, UPC
MPI, OpenMP, shmem
Totalview debugger, variety of performance tools

Libraries

BLAS, LAPACK, ScLAPACK, ESSL, PSSL, scilib, TOPS

FY06 Allocations



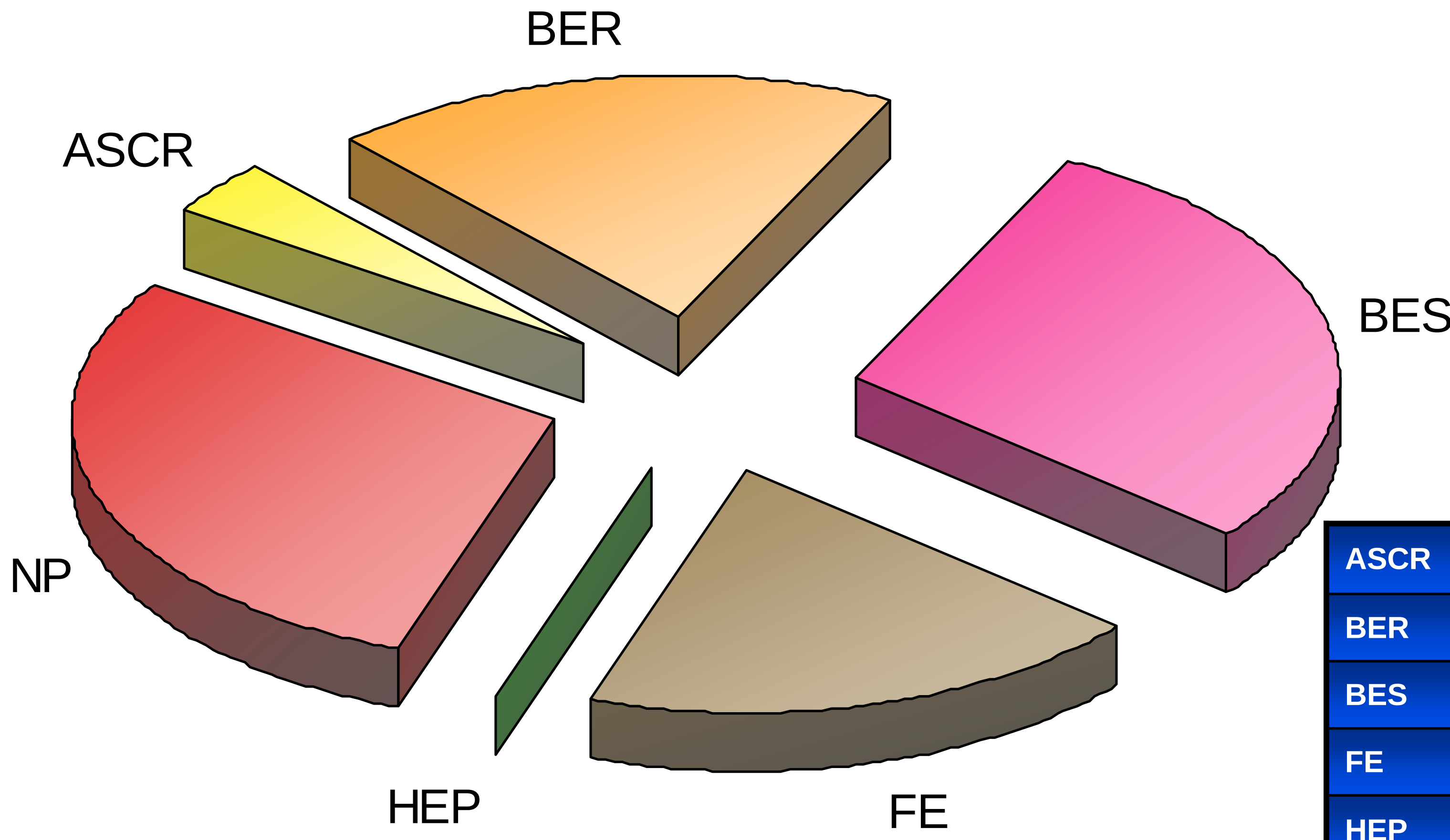


Projects in areas of

- Accelerator design
- Astrophysics
- Biostructures
- Catalysis
- Climate
- Combustion
- Fusion
- Materials
- Ocean turbulence

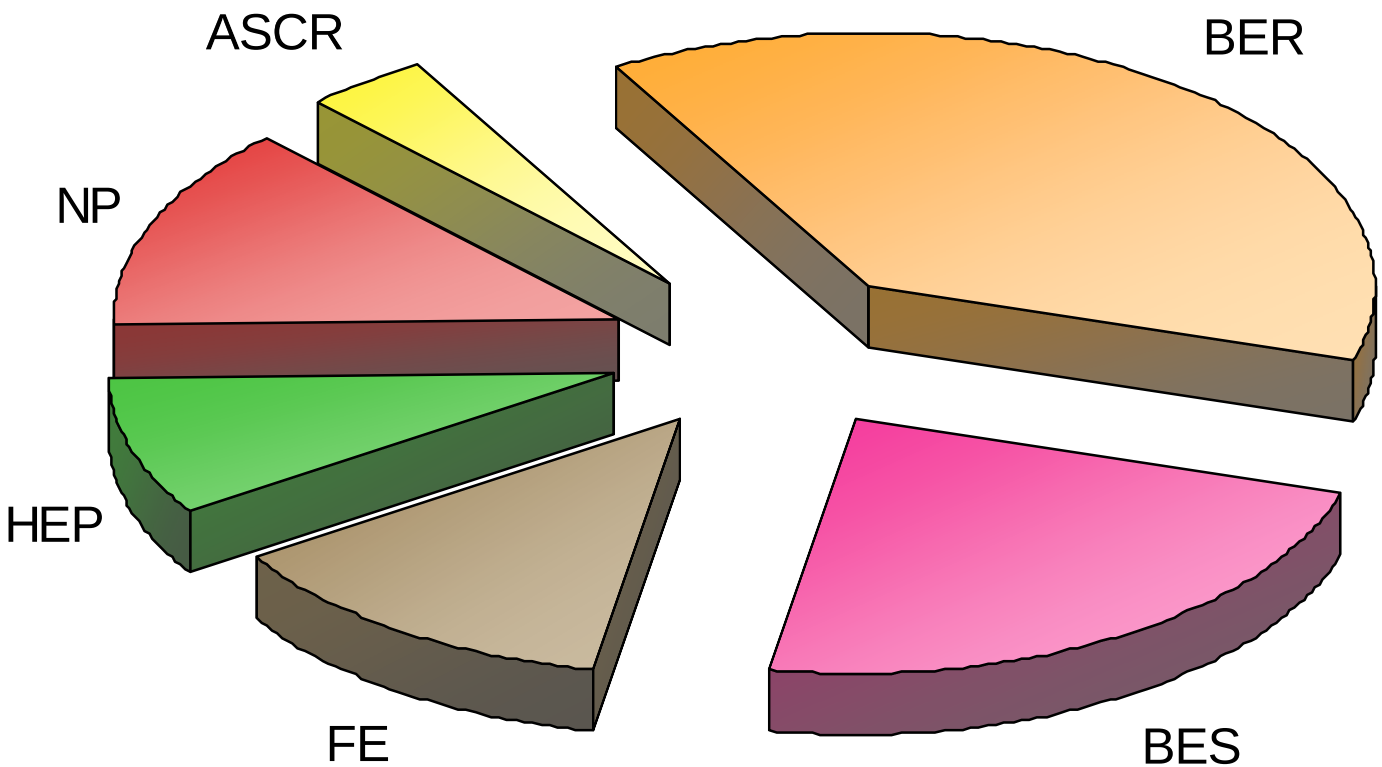
Contact: help@nccs.gov

Total Proposed Cray XT3 Allocations per Program Office



ASCR	1,000,000	4%
BER	4,996,856	19%
BES	7,500,000	29%
FE	5,000,000	19%
HEP	30,000	<1%
NP	7,550,000	29%
Total	26,076,856	100%

Total Proposed Cray X1E Allocations per Program Office



ASCR	200,000	4%
BER	2,029,000	38%
BES	1,200,000	23%
FE	665,240	13%
HEP	500,000	9%
NP	700,000	13%
Total	26,076,856	100%

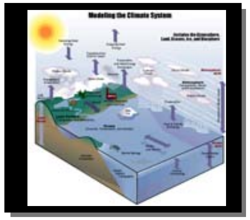
Current ASCR Allocations

Performance Evaluation and Analysis Consortium (PEAC) End Station

Patrick Worley, ORNL

Cray XT3 Jaguar: 1,000,000 processor hours and Cray X1E Phoenix: 200,000 processor hours

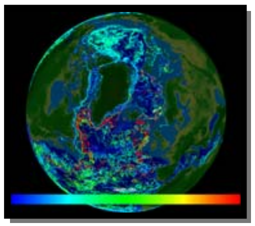
Current BER Allocations



Climate-Science Computational End Station Development and Grand Challenge Team

Warren Washington, NCAR

Cray XT3 Jaguar: 3,000,000 processor hours and Cray X1E Phoenix: 2,000,000 processor hours



Eulerian and Lagrangian Studies of Turbulent Transport in the Global Ocean

Synte Peacock, Univ. of Chicago

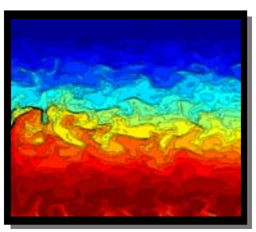
Cray XT3 Jaguar: 1,496,856 processor hours

Next Generation Simulations in Biology:

Investigating Biomolecular Structure, Dynamics and Function Through Multi-Scale Modeling

Pratul Agarwal, ORNL

Cray XT3 Jaguar: 500,000 processor hours

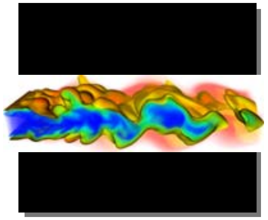


The Role of Eddies in the Thermohaline Circulation

Paola Cessi, Scripps Institution of Oceanography, UCSD, CA

Cray X1E Phoenix: 29,000 processor hours

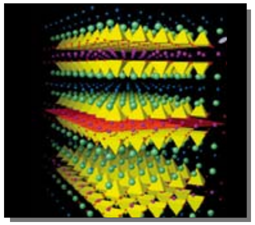
Current BES Allocations



High-Fidelity Numerical Simulations of Turbulent Combustion - Fundamental Science Towards Predictive Models

Jackie Chen, SNL

Cray XT3 Jaguar: 3,000,000 processor hours and Cray X1E Phoenix: 600,000 processor hours



Predictive Simulations in Strongly Correlated Electron Systems and Functional Nanostructures

Thomas Schulthess, ORNL

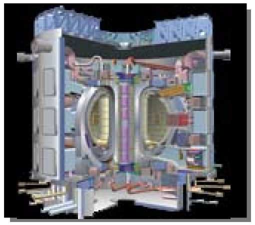
Cray XT3 Jaguar: 3,500,000 processor hours and Cray X1E Phoenix: 300,000 processor hours

Current FE Allocations

Exploring Advanced Tokamak Operating Regimes Using Comprehensive GYRO Gyrokinetic Simulations

Jeff Candy, General Atomics

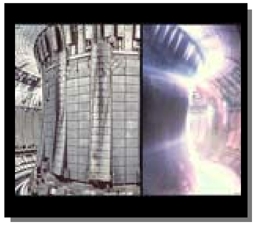
Cray X1E Phoenix: 440,240 processor hours



Gyrokinetic Plasma Simulation

W. W. Lee, PPPL

Cray XT3 Jaguar: 2,000,000 processor hours and Cray X1E Phoenix: 225,000 processor hours



Simulation of Wave-Plasma Interaction and Extended MHD in Fusion Systems

Don Batchelor, ORNL

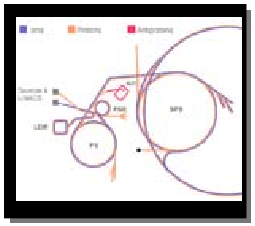
Cray XT3 Jaguar: 3,000,000 processor hours

Current HEP Allocations

Computational Design of the Low-loss Accelerating Cavity for the ILC

Kwok Ko, SLAC

Cray X1E Phoenix: 500,000 processor hours

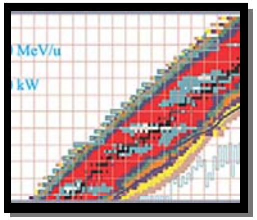


Monte Carlo Simulation and Reconstruction of CompHEP–Produced Hadronic Backgrounds to the Higgs Boson Diphoton Decay in Weak-Boson Fusion Production Mode

Harvey Newman, California Institute of Technology

Cray XT3 Jaguar: 30,000 processor hours

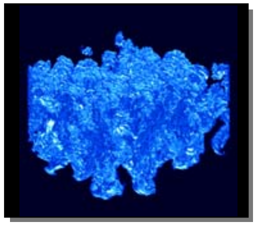
Current NP Allocations



Ab-Initio Nuclear Structure Computations

David Dean, ORNL

Cray XT3 Jaguar: 1,000,000 processor hours



Ignition and Flame Propagation in Type Ia Supernovae

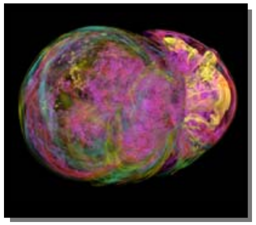
Stan Woosley, UC Santa Cruz

Cray XT3 Jaguar: 3,000,000 processor hours

Multi-Dimensional Simulations of Core-Collapse Supernovae

Adam Burrows, University of Arizona

Cray XT3 Jaguar: 1,250,000 processor hours



Multi-Dimensional Simulations of Core-Collapse Supernovae

Anthony Mezzacappa, ORNL

Cray XT3 Jaguar: 3,550,000 processor hours and Cray X1E Phoenix: 700,000 processor hours

Current INCITE Allocations

Development and Correlations of Large Scale Computational Tools for Flight Vehicles

Moeljo Hong, The Boeing Company

Cray X1E Phoenix: 200,000 processor hours

Direct Numerical Simulation of Fracture, Fragmentation, and Localization in Brittle and Ductile Materials

Michael Ortiz, California Institute of Technology

Cray XT3 Jaguar: 500,000 processor hours

Interaction of ETG and ITG/TEM Gyrokinetic Turbulence

Ronald Waltz, General Atomics

Cray X1E Phoenix: 400,000 processor hours

Molecular Dynamics Simulations of Molecular Motors

Martin Karplus, Harvard University

Cray XT3 Jaguar: 1,484,800 processor hours

Real-Time Ray-Tracing

Evan Smyth, Dreamworks

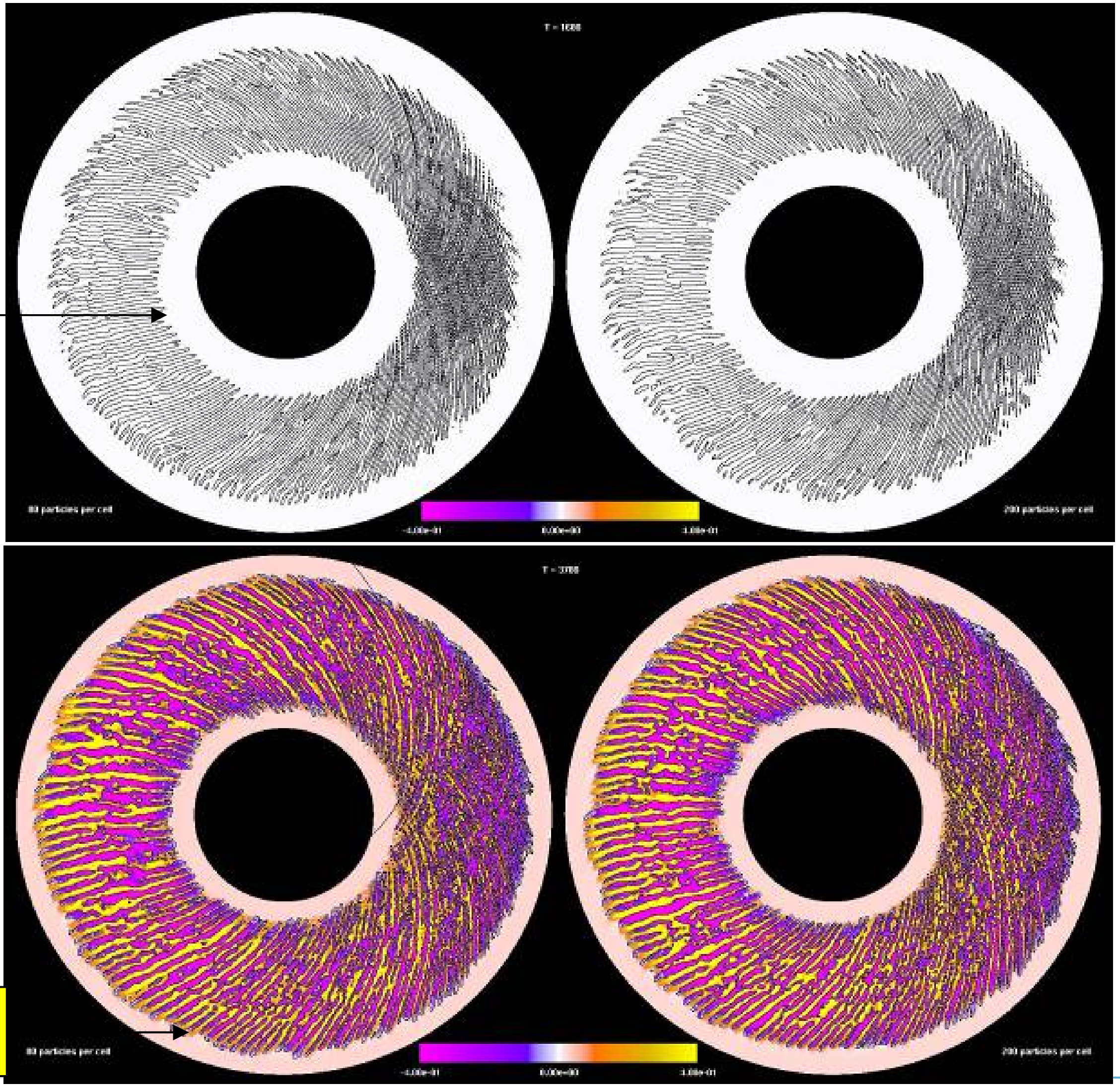
Cray XT3 Jaguar: 950,000 processor hours

An Example of Recent Results: Fusion (W. Lee – PPPL)

- Record size simulations on XT3 shows convergence for ETG runs

- **<u>20 billion particles!</u>**
- **4,800 processors**
- **200 particles/cell**
- **Made possible by ORNL NCCS computer**

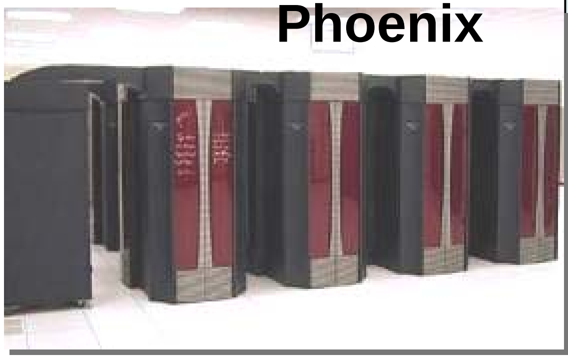
- **~ 10 TBytes of memory!**



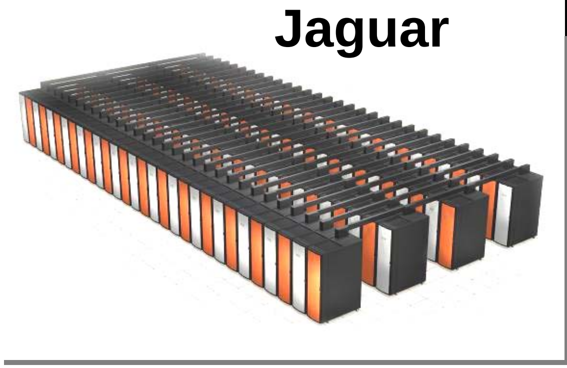
Site-Wide High-Performance File System

- To serve all NCCS resources
- Cost effective solution
  - **Decouple supercomputer procurements from storage**
  - **Easier and high-performance data migration between NCCS resources**
- Spider
  - **Prototype proof-of-the-concept testbed**
  - **20 OSS, 1 MDS, dual-socket dual-core AMD Opterons**
  - **2 DDN 8500 couplets, 1 DDN 9500**
  - **10 Gbps Ethernet, 4X IB, and Myrinet 10 Gbps**
  - **Lustre file system**
  - **~2.5 GB/s throughput obtained**
- Target
  - **10 GB/s by the end of ’06**
  - **60 GB/s by the end of ‘07**

Two application-driven architectures are converging as Cascade

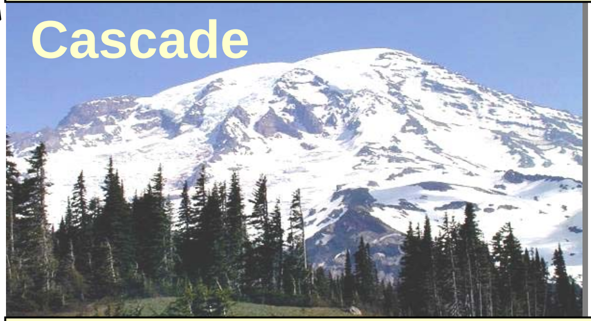


- Most powerful processors and interconnect
- Scalable, globally addressable memory and bandwidth



- Extremely low latency, high bandwidth interconnect
- Efficient scalar processors, balanced interconnect

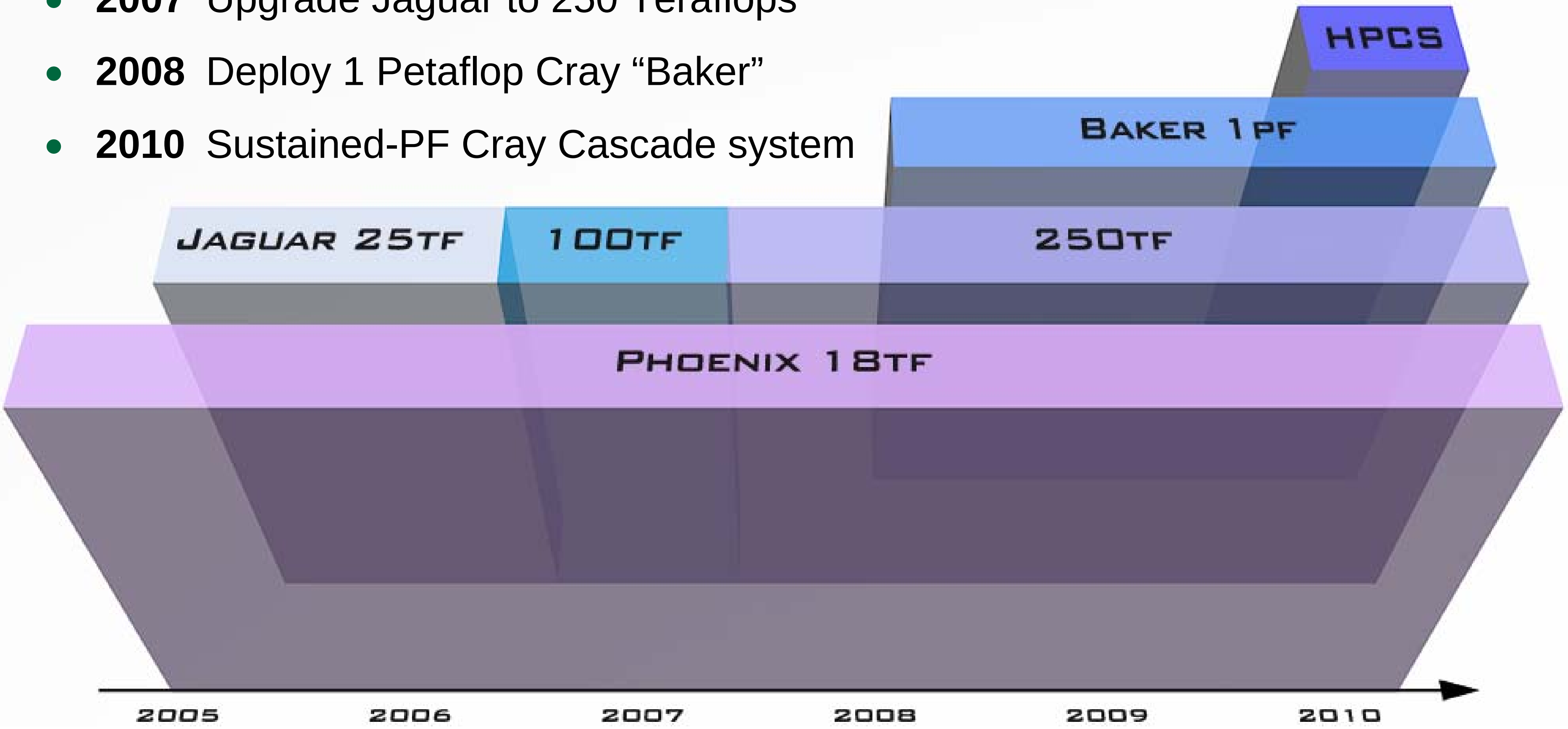
- Unified system including vector, scalar, multithreaded and potentially FPGA processors
- Scalable network and globally addressable memory
- Adaptive custom processors



- Single Linux-based user interface and environment
- Shared global file system
- Improved performance by matching processor to job
- Single solution for diverse workloads

Hardware Roadmap

- **Currently in production:** 18 TF Phoenix and 25 TF Jaguar
- **2006** Upgrade Jaguar to 100 Teraflops
- **2007** Upgrade Jaguar to 250 Teraflops
- **2008** Deploy 1 Petaflop Cray "Baker"
- **2010** Sustained-PF Cray Cascade system



Questions?

NATIONAL CENTER
FOR COMPUTATIONAL SCIENCES

presented by
Sarp Oral, PhD
oralhs@ornl.gov